Title
The Role of Information in Building Reputation in an Investment/Trust Game

Journal
European Accounting Review, 22(3)

ISSN
0963-8180

Author
Lunawat, R

Publication Date
2013-09-01

DOI
10.1080/09638180.2012.748256

Peer reviewed

The Role of Information in Building Reputation in an Investment / Trust Game[*, **]

Radhika Lunawat[†]

November 2012

* This paper is based on my PhD dissertation at the University of Minnesota. I am indebted to my dissertation committee for valuable suggestions and guidance: John Dickhaut (Chair), Beth Allen, Chandra Kanodia, and Gregory Waymire. I wish to thank Ron King, Christian Leuz, Harry Evans, Jack Stecher, Regina Anctil, associate editor John Christensen, and two anonymous referees for helpful comments and suggestions. I also wish to thank workshop participants at the London School of Economics, the Indian School of Business, the University of Pittsburgh, Argyros School of Business and Economics (Chapman University), the University of Toronto, the University of Alberta, and the Economic Science Institute (Chapman University); and referees for the National Science Foundation Doctoral Dissertation Grant and for the University of Minnesota Graduate School Fellowship; and various conference participants.

**I gratefully acknowledge support from National Science Foundation Doctoral Dissertation Grant Number 0820455 and from Accounting Research Center at Carlson School of Management, University of Minnesota.

[†]Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA – 15213; e-mail: rlunawat@andrew.cmu.edu *and* Opus School of Business, University of St Thomas, 1000 LaSalle Avenue, Minneapolis, MN – 55403; e-mail: rlunawat@stthomas.edu.

Abstract

This article analyzes the role of information in building reputation in an investment/trust game. The model allows for information asymmetry in a finitely repeated sender–receiver game and solves for sequential equilibrium to show that if there are some trustworthy managers who always disclose their private information and choose to return a fair proportion of the firm's income as dividend to the investor, then a rational manager will mimic such behaviour in an attempt to earn a reputation for being trustworthy. The rational manager will mimic with probability 1 in the early periods of the game. The investor, too, will invest with probability 1 in these periods. However, in the later periods, the rational manager will mimic with a certain probability strictly less than 1. The probability will be such that it will make the investor indifferent between investing and not investing, and he, in turn, will invest with a probability (strictly less than 1) that will make the rational manager indifferent between mimicking and not mimicking; that is, the game will begin with pure strategy play but will switch to mixed strategy play. There is one exception, though: when the investor's ex ante beliefs about the manager's trustworthiness are exceptionally high, the game will continue in a pure strategy, and the switch to mixed strategy play will never occur. Identical results obtain if the manager's choice of whether to share his private information with the investor is replaced by exogenously imposed information sharing.

Keywords: Information, Disclosure, Reputation, Trust

JEL codes: M41, D82, C11, C73.

1. Introduction

> In the United States, home to some of the largest corporate collapses, trust in business collapsed as well, dropping 20 percentage points over the course of one year. With only 38% of informed public in the United States trusting business today, levels are the lowest they have been in the Barometer's tracking history—even lower than in the wake of Enron and the dot-com bust. . . . For U.S. businesses, this downturn marks a stark reversal from the steady uptick in trust of the last five years.
>
> —Edelman Trust Barometer, 2009

In the face of the financial meltdown, one of the challenges managers face is to rebuild their reputation for being trustworthy. This article analyzes the role of information in facilitating such investor–manager trust and reputation in an investment/trust game.

The investment/trust game (Berg *et al.*, 1995) is a sender–receiver game in which the sender/investor is endowed with some wealth and chooses whether to invest it in a receiver/manager. The receiver is endowed with some production technology to which he channels the investment for productive uses, thereby generating a strictly positive income for the firm comprising of the investor and the manager. From the firm's income, the manager chooses whether to pay a dividend back to the investor and then keeps the residual for himself. When this game is repeated for some periods (Dickhaut *et al.*, forthcoming), it allows the managers an opportunity to build a reputation for being trustworthy. Fundamental to such reputation building is the existence of a trustworthy manager, which is defined as a manager who pays back a fair proportion of the firm's income as dividend to the investor. The rational manager tries to build a reputation for being trustworthy by mimicking such behaviour in the early periods of a repeated investment game.

The only source of uncertainty in the repeated investment game is the manager's type, and in the early periods of the game, it is partially resolved by the manger's observable action of paying back the dividend. Exploring the role of information in such a setting necessitates the introduction of an additional source of uncertainty, namely, uncertainty regarding the state of nature (Dickhaut *et al*., 2012). The deterministic income of the firm is now replaced by one that is contingent on the state of nature. Depending on the state of nature that obtains, the firm's income may be high or low. The manager always observes the state of nature, but the investor does not. This information asymmetry implies that a low dividend could mean one of the following two things for an investor: (1) it could be that the firm's income is low or (2) it could be that the firm's income is high but a rational manager is trying to pretend that a low income obtained. Information sharing with the investor becomes important in such a setting and has potential implications for managerial efforts at building reputation for trustworthiness. This article analyzes such a reputation-building role of information.

When information sharing is exogenously imposed and alleviates information asymmetry, the previously mentioned definition of a trustworthy manager suffices. However, if information sharing with the investor is voluntary, then a consideration of reputation building necessitates changing the definition of a trustworthy manager to one who pays back a fair proportion of the firm's income as dividend to the investor *and* voluntarily discloses his private information about the state of nature. Mimicking a trustworthy type disciplines the actions of a rational manager in that it takes away from him the option of paying back a low dividend (and thereby pretending that a low state of nature has occurred) when the high state of nature has occurred. Whether information sharing is exogenously imposed or is a voluntary choice of the manager, what is important is the disciplining role it has for a rational manager.

From a reputation building standpoint, such perfect mimicking and concomitant discipline are optimal for a rational manager in the early periods of a finitely repeated game. This is so because in the early periods, the expected future benefits from mimicking exceed the costs of such mimicking. The benefits to the manager flow from future investments of the investor. The cost arises from having to pay back some portion of the income of the firm as dividend to the investor. In the early periods, the benefit of expected future investments exceeds the cost of dividend in the current period, making perfect mimicking optimal. The investor is a Bayesian updater, and perfect mimicking by the rational manager implies that the investor's posterior beliefs about a manager's trustworthiness are the same as his prior beliefs about the manager's trustworthiness. The investor chooses to invest if his beliefs about the manager's trustworthiness are above the threshold specified by the model. This threshold increases over time because the investor knows that the rational manager has a strong incentive to mimic the trustworthy type and pay a fair dividend in the early periods. On one hand, the investment threshold increases over time, but on the other hand, perfect mimicking implies that the investor's beliefs about the manager's trustworthiness do not change. The model specifies the point in the repeated game at which the beliefs sustained by perfect mimicking will be below the investment threshold.

To prevent beliefs from falling below the investment threshold, the rational manager switches from a strategy of perfect mimicking to a strategy of selective mimicking; that is, he mimics the trustworthy manager with a certain probability (specified by the model) strictly less than 1. The probability is such that the investor's updated beliefs about the manager's trustworthiness are exactly on the investment threshold. Note that the investor invests with certainty only when his beliefs are above the investment threshold. However, when his beliefs are exactly on the investment threshold, he is indifferent between investing and not investing. He chooses to invest with a certain probability (specified by the model) that makes the rational manager indifferent between mimicking and not mimicking.

This article is closely related to a couple of recent papers that analyze reputation and disclosure. Einhorn and Ziv (2008) analyze the reputation of a firm for being informationally endowed and how disclosure impacts such reputation. Beyer and Dye (forthcoming) look at how a manager builds reputation for being a type that shares his private information. They cast it as a manager's decision-making problem and ensure consistency with a pricing function. I cast it in a game theoretic framework, in which both the (rational) manager and the investor are strategic players who act in their best interest. In the Beyer and Dye model, there is uncertainty about the manager's type and about his information endowment, allowing for studying the effects the interaction of the two sources of uncertainty have on reputation building. In my model, the manager always receives private information. The uncertainty here is about the manager's type and the state of nature. Furthermore, the Beyer and Dye model has a continuum of manager's types, and this allows for equilibrium in pure strategies. In contrast, types in my model are binary (namely, trustworthy and untrustworthy/rational), and equilibrium allows for mixed strategy play.

This approach of binary types, uncertainty about types, and possibility of mixed strategies in equilibrium follows from reputation models in economics. For example, Kreps and Wilson (1982) solve a sequential equilibrium for a finitely repeated monopolist entry deterrence game in which the monopolist is either strong or weak and the uncertainty pertains to the type of the monopolist. Similarly, Kreps *et al.* (1982) solve for a sequential equilibrium in a finitely repeated prisoners' dilemma game. I solve for a sequential equilibrium in my model, but relative to these reputation models, I have an additional source of uncertainty: the uncertainty about the state of nature.

The rest of the article proceeds as follows. Section 2 introduces the model. Section 3 defines and characterizes the equilibrium of the model. Section 4 discusses the properties of the equilibrium. Section 5 summarizes and concludes.

2. Model

There are two players, a sender/investor and a receiver/manager. Nature moves first and selects the manager's type to be either trustworthy or untrustworthy. I will define momentarily what I mean by each type. The manager knows his type, but the investor does not. The game then proceeds through n periods, in each of which the investor and the manager make a sequence of choices. In what follows, the subscript t ($t = 1, 2, \ldots, n$) will be used to denote a period. The manager chooses whether to commit to truthfully disclosing private information he will learn over the course of the game. A decision to commit to such truthful disclosure is denoted by $d_t = 1$, and a decision not to commit to such truthful disclosure is denoted by $d_t = 0$. Note that the manager is not privy to the private information at the time he makes the choice of whether to commit to truthfully disclosing it; rather, it is information he *will* learn over the course of the game. It is as if the manager is making a choice of an accounting system—the manager could choose an accounting system that will generate information that both the investor and the manager will learn (by choosing to commit to truthfully disclosing), or alternatively, the manager could choose an accounting system that will generate information only the manager will learn (by choosing not to commit to truthfully disclosing). Note that if the manager were instead to make a decision on truthfully disclosing his private information after he learns it, then such a decision would predicate on the information content. For example, if the private information is good, he may choose to disclose it, but if it is bad, he may choose to withhold it. The ex ante commitment modelled in this article obviates commingling managerial strategic considerations arising from the information content with the strategic considerations arising from innate building of reputation for being trustworthy and allows focus on the latter.

The investor sees whether the manager has committed to truthfully disclosing his private information. He is endowed with $e > 0$ units of wealth and chooses whether to invest in the manager. Regardless of whether the investor chooses to invest, e is common knowledge; that is, both the investor and the manager know the amount of wealth with which the investor is endowed. The investor's decision to invest is denoted by $m_t = e$, and the investor's decision not to invest is denoted by $m_t = 0$. If the investor chooses to invest, then the amount e is multiplied by a multiplier λ_t before the manager receives it; $\lambda_t \in \{l, h\}$ and is equally likely to be either l or h in every period: $1 \leq l < 2 < h$, $l + h > 4$. It is as if the manager has some production technology because of which he is able to grow the investment of e to $e\lambda_t$. The multiplied amount ($e\lambda_t$) may be thought of as the gross income of the firm comprising the investor and the manager. I will explain momentarily the parameter restriction of $1 \leq l < 2 < h$, $l + h > 4$.

Now the manager receives $e\lambda_t$ and learns λ_t. However, the investor learns λ_t only if the manager had earlier committed to truthfully disclosing his private information; that is, if the manager had chosen an accounting system that generates information both the investor and the manager learn, then the investor learns λ_t. Otherwise, if the manager had chosen an accounting system that generates information only the manager learns, then the investor does not learn λ_t. In this sense, λ_t is the manager's private information—he always learns the realized value of λ_t, but the investor's knowledge of λ_t is dependent on the manager's choice of accounting system.

After the manager receives $e\lambda_t$, he chooses to send back k_t to the investor. If $\lambda_t = l$, $k_t \in \{0, el/2\}$; if $\lambda_t = h$, $k_t \in \{0, el/2, eh/2\}$. The idea is that if the low state of the world (namely, l)

occurs, then the manager either can choose to send back half or choose to send back nothing to the investor. However, if the high state of the world occurs (namely, h), then the manager has the additional option of choosing to send back an amount (namely, $el/2$) consistent with the low state of the world having occurred[1]. The investor receives k_t, and the manager keeps the residual $e\lambda_t - k_t$. The amount sent back by the manager (k_t) may be thought of as the dividend the manager pays to the investor. A trustworthy manager is defined as one who always chooses to disclose ($d_t = 1$) and always chooses to return half of what he receives (i.e., if $\lambda_t = l$, he chooses $k_t = el/2$, and if $\lambda_t = h$, he chooses $k_t = eh/2$). An untrustworthy manager is defined as a manager who is not trustworthy; in other words, the untrustworthy manager is a rational manager whose action is guided by self-interest. I have $l/2 < 1$ so that a realization $\lambda_t = l$ implies a negative net return for the investor, $h/2 > 1$ so that a realization $\lambda_t = h$ implies a positive net return for the investor, and $(l + h)/4 > 1$ so that the expected net return for the investor, if the manager is trustworthy, is positive. Risk neutrality, additively separable utility, and no time discounting are assumed.

[1]Note that this idea can also be captured by expanding the set of k_t to include more elements than the ones specified here. The equilibrium and other results derived will be qualitatively similar. The set of k_t used here is the most parsimonious one possible.

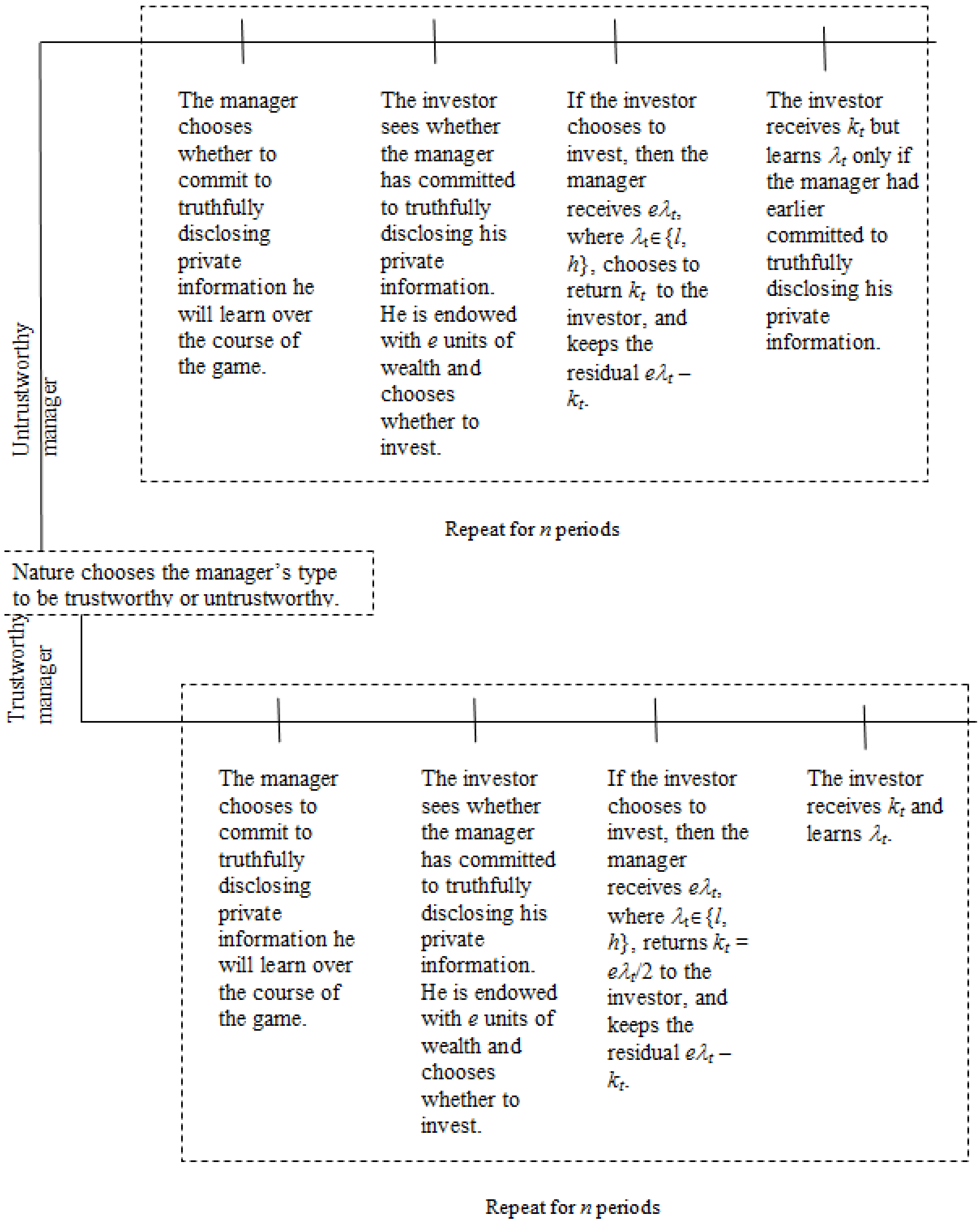

***Figure 1**. Timeline of the model*

3. Equilibrium

Because this is a finitely repeated game with two players and uncertainty about the type of one of the players, I can solve for either a sequential equilibrium or a perfect Bayesian equilibrium. Kreps and Wilson (1982) and Kreps *et al.* (1982) are examples of papers that solve for a sequential equilibrium in such reputation games, whereas Dickhaut *et al.* (forthcoming) is an example of a paper that solves for a perfect Bayesian equilibrium. Either solution concept will yield identical results in this article because Fudenberg and Tirole (1991) show that if each player has at most two possible types (the manager can be either trustworthy or untrustworthy/rational in the model in this article), then the sets of perfect Bayesian equilibria and sequential equilibria coincide.

A sequential equilibrium of the game described in Section 2 is defined as follows. An equilibrium comprises a strategy for each player and, for each period *t*, a function P_t^D that takes the history of moves up to period *t* into numbers in [0 , 1] such that

1. Starting from any point in the game where it is the manager's move, the manager's strategy is a best response to the investor's strategy.
2. Starting from any point in the game where it is the investor's move, the investor's strategy is a best response to the manager's strategy, given that the investor believes that the manager is trustworthy with probability $P_t^D(h_t)$.
3. The game begins with $P_0^D = \delta$.
4. Each P_t^D is computed from P_{t-1}^D and the manager's strategy using Bayes's rule whenever possible.

I will first define the function P_t^D. However, before I define P_t^D, note that the set of subscripts for P_t^D is different from the set of subscripts for each of d_t, m_t, λ_t, and k_t. While each of d_t, m_t, λ_t, and k_t is subscripted using the set $\{1, 2, \ldots, n\}$, P_t^D is subscripted using the set $\{0, 1, 2, \ldots, n\}$. The investor enters the game with $P_0^D = \delta$. He sees d_1 and updates to P_1^D before choosing m_1. Because the investor's updating from P_0^D to P_1^D occurs before his choice of investment for the first period (m_1), this necessitates the introduction of an additional element (namely, 0) in the set of subscripts for P_t^D.

Set $P_0^D = \delta$. Then, P_t^D may be defined as follows:

P(i) If the manager does not disclose in period 1 (i.e., the manager chooses $d_1 = 0$), then $P_1^D = 0$, and if the manager discloses in period 1 (i.e., the manager chooses $d_1 = 1$), then $P_1^D = P_0^D = \delta$.

P(ii) For $t > 1$, if the investor does not invest in period $t - 1$ (i.e., the investor chooses $m_{t-1} = 0$), then $P_t^D = P_{t-1}^D$.

P(iii) For $t > 1$, if the investor invests in period $t - 1$ (i.e., the investor chooses $m_{t-1} = e$) and the manager discloses in period t, and the manager returns half of what he receives in period $t - 1$ and $P_{t-1}^D > 0$ (i.e., the manager chooses $d_t = 1$ and $k_{t-1} = e\lambda_{t-1}/2$), then $P_t^D = \max\left((4/(l+h))^{n-t+1}, \delta\right)$.

P(iv) For $t > 1$, if the investor invests in period $t - 1$ (i.e., the investor chooses $m_{t-1} = e$) but the manager either does not disclose in period t or does not return half of what he receives in period $t - 1$ (i.e., the manager chooses either $d_t = 0$ or $k_{t-1} = 0$), then $P_t^D = 0$.

P(v) If $P_{t-1}^D = 0$, then $P_t^D = 0$.

P_t^D is the probability with which the investor believes the manager is trustworthy. The game begins with $P_0^D = \delta$. If the manager does not disclose in period 1, then the investor knows that the manager is untrustworthy and therefore sets $P_1^D = 0$ (point P(i)). For $t > 1$, if the history of play up to period t either includes any instance in which the manager fails to disclose or includes any instance in which the manager fails to return half of what he receives, then the investor knows that the manager is untrustworthy and therefore sets $P_t^D = 0$ (point P(iv)). If the investor does not invest in any given period, then no Bayesian updating occurs, and the investor's posterior beliefs about the manager's trustworthiness are the same as his prior beliefs about the manager's trustworthiness (point P(ii)). If the manager has always chosen to disclose and has always returned half of what he receives, then $P_t^D = \max\left((2/E(\lambda_t))^{n-t+1}, \delta\right)$ (point P(iv)). The question is, if the manager has always chosen to disclose and has always returned half of what he receives, then when is it that $P_t^D = (2/E(\lambda_t))^{n-t+1}$, and when is it that $P_t^D = \delta$? To answer this, note first that a trustworthy manager, by definition, always chooses to disclose and always chooses to return half of what he receives. I will argue later that the rational manager mimics such behaviour of the trustworthy manager to develop a reputation for being trustworthy. If the rational manager does such mimicking with certainty, then no Bayesian updating occurs, and the investor's posterior beliefs about the manager's trustworthiness are the same as his prior beliefs about the manager's trustworthiness; that is, $P_t^D = P_{t-1}^D = \ldots = P_0^D = \delta$. However, if the rational manager does such mimicking with a certain probability strictly less than 1, then the investor's updated beliefs about a manager's trustworthiness are given by $P_t^D = (4/(l+h))^{n-t+1}$.

In the proof in the appendix, I show that this is consistent with the Bayesian updating criterion specified in the definition of the equilibrium.

Now, I will describe the strategies of the investor and the untrustworthy manager in terms of P_t^D. The investor's strategy may be outlined as follows:

I(i) If $P_t^D < (4/(l+h))^{n-t+1}$, the investor chooses not to invest, that is, $m_t = 0$.

I(ii) If $P_t^D > (4/(l+h))^{n-t+1}$, the investor chooses to invest, that is, $m_t = e$.

I(iii) If $P_t^D = (4/(l+h))^{n-t+1}$, with a probability $V_t^D = \lambda_{t-1}/(l+h)$, the investor chooses to invest, and with a probability $1-V_t^D$, the investor chooses not to invest. That is, with a probability V_t^D, the investor chooses $m_t = e$, and with a probability $1-V_t^D$, the investor chooses $m_t = 0$.

The investor follows a threshold strategy in which he invests if his beliefs about the manager's trustworthiness (P_t^D) are above the threshold for the period ($(4/(l+h))^{n-t+1}$) (point I(ii)) and does not invest if his beliefs are below the threshold (point I(i)). When his beliefs are on the investment threshold (point I(iii)), he is indifferent between investing and not investing. He chooses to invest with a probability V_t^D that makes the rational manager indifferent between mimicking (to be a trustworthy type) and not mimicking.

By definition, a trustworthy manager always chooses to disclose and always chooses to return half of what he receives. An untrustworthy/rational manager's strategy depends on t and P_t^D. The strategy may be outlined as follows:

M(i) $d_1 = 1$.

M(ii) If $t = n$, the manager does not return anything; that is, the manager chooses $k_n = 0$.

M(iii) If $t < n$ and $P_t^D \geq (4/(l+h))^{n-t}$, the manager chooses to disclose and chooses to return half of what he receives; that is, the manager chooses $d_{t+1} = 1$ and $k_t = e\lambda_t / 2$.

M(iv) If $t < n$ and $P_t^D < (4/(l+h))^{n-t}$, then with a probability $S_t^D = P_t^D(1-(4/(l+h))^{n-t})/(4/(l+h))^{n-t}(1-P_t^D)$, the manager chooses to return half of what he receives in period *t*. With probability $1-S_t^D$, the manager chooses not to return anything in period *t*. In the instance in which the manager returns half of what he receives in period *t*, he chooses to disclose in period *t +1*; that is, with a probability S_t^D, the manager chooses $d_{t+1} = 1$ and $k_t = e\lambda_t / 2$, and with a probability $1-S_t^D$, the manager chooses $k_t = 0$. Note that if $P_t^D = 0$, then $S_t^D = 0$, and if $P_t^D = (4/(l+h))^{n-t}$, then $S_t^D = 1$.

The untrustworthy/rational manager always chooses to disclose in period 1 (point M(i)) because not doing so reveals to the investor that the manager is untrustworthy. He chooses not to return anything as dividend in the last period (point M(ii)) because the alternative of paying a nonzero dividend is personally costly to him, and because it is the last period, there are no associated expected future benefits. In periods before the last one, he mimics the trustworthy manager (points M(iii) and M(iv)) and therefore chooses to disclose his private information and to pay back half of the firm's income as dividend to the investor. The rational manager does such mimicking with certainty until a certain point in the game (point M(iii)) and then switches to mimicking with a certain probability (S_t^D) strictly less than 1 (point M(iv)). The intuition behind why the rational manager makes this switch is explained next.

From I(i)–I(iii), the investor invests in period *t + 1* if his beliefs about the manager's trustworthiness are above the investment threshold for that period, that is, if $P_{t+1}^D \geq (4/(l+h))^{n-t}$. Consider $P_t^D \geq (4/(l+h))^{n-t}$. If the manager mimics with certainty (point M(iii)), then no Bayesian updating occurs, and hence $P_{t+1}^D = P_t^D \geq (4/(l+h))^{n-t}$. Because P_{t+1}^D is above the investment threshold for period *t + 1*, the investor will invest. Now, consider $P_t^D < (4/(l+h))^{n-t}$. If the manager continues to mimic with certainty, then $P_{t+1}^D = P_t^D < (4/(l+h))^{n-t}$, and the investor will not invest in period *t + 1*. Therefore, the manager switches to mimicking with probability $S_t^D < 1$ (point M(iv)). In the proof in the appendix, I show that S_t^D is so chosen that the investor's updated period *t + 1* beliefs about the manager's trustworthiness are exactly on the investment threshold for period *t + 1*. From point I(iii), the investor chooses to invest with probability $V_{t+1}^D < 1$ when his beliefs are exactly on the threshold.

Summarizing, the game starts with pure strategy play in which the investor invests with probability 1 and the rational manager mimics with probability 1. But at some point in the game, it switches to mixed strategy play in which the investor invests with probability $V_{t+1}^D < 1$ and the rational manager mimics with probability $S_t^D < 1$. There is one exception: when the investor's period zero beliefs about a manager's trustworthiness are above the investment threshold for the last period (i.e., $P_0^D = \delta > 4/(l+h)$), then the game continues in pure strategies and the switch to mixed strategy play never occurs.

Proposition. The strategies (I(i)–I(iii), M(i)–M(iv)) and beliefs (P(i)–P(v)) described above constitute a sequential equilibrium.

The proof is outlined in the appendix.

4. Properties of the Equilibrium

4.1. Increasing Investment Threshold

P_{t+1}^{D} must be at least $(4/(l+h))^{n-t+1}$ for the investor to invest with some positive probability. This investment threshold $(4/(l+h))^{n-t+1}$ may also be thought of as the manager's reputation. The manager starts mixed strategy play in some period *t* to ensure that the investor's updated (period *t* + *1*) beliefs about the manager's type are exactly on the threshold. This threshold increases in *t*; that is, the manager's reputation over time must be progressively higher for the investor to invest. For example, if *n* = 10, *l* = 1, and *h* = 5, then this investment threshold may be graphed as in Figure 2A.

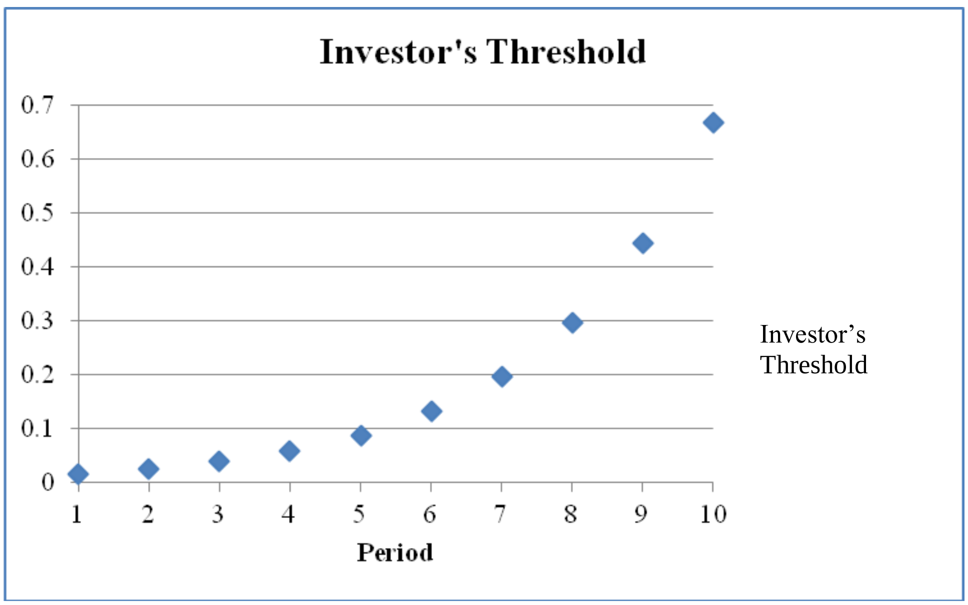

Figure 2A*. Investment threshold at which the investor invests with some probability when n = 10, l = 1, and h = 5*

The idea is that in the early periods of a repeated game, the investor knows that the rational manager is extremely likely to mimic with certainty, and this implies that it is extremely likely the investor gets a fair dividend. Therefore he invests even when his beliefs about the manager's trustworthiness are low; that is, the investment threshold is low in the early periods. However, in later periods, mixed strategy play may start, and then the rational manager is not as likely to pay out a fair dividend. Therefore the threshold at which the investor invests is now higher.

4.2. Perfect Mimicking Strategy

What happens if the untrustworthy/rational manager follows a strategy of choosing $d_t = 1$ and $k_t = e\lambda_t / 2$? That is, the untrustworthy manager chooses to mimic to be the trustworthy type with probability 1. This strategy will be a part of the equilibrium described earlier if $P_0^D \geq 4/(l+h)$. However, when $P_0^D < 4/(l+h)$, then by Bayesian updating, $P_{t+1}^D = P_t^D = ... = P_0^D$. At some t = i, $P_i^D < \left(4/(l+h)\right)^{n-i+1}$, and therefore the investor does not invest in any period $t \geq i$. In contrast, if the manager follows the strategy outlined in the equilibrium described earlier, the investor invests with some positive probability in period $t \geq i$. In other words, perfect mimicking does not constitute equilibrium and will result in lower investment. In equilibrium, the manager mimics to be the trustworthy type but does so selectively instead of indiscriminately to ensure the credibility associated with his choice in the instances in which he actually chooses to mimic (except when $P_0^D \geq 4/(l+h)$).

For example, if n = 10, l = 1, h = 5, and $P_0^D = 0.4$, then the investment threshold is given by the blue dotted line in Figure 2B. The way the investor's beliefs about the manager's type evolve in equilibrium and under the perfect mimicking strategy, respectively, are shown by the red dotted line and the green dotted line in Figure 2B. Note that under the perfect mimicking strategy, the investor's beliefs fall below the investment threshold in period 9 but remain on the threshold under equilibrium. Consequently, under perfect mimicking, the investor does not invest in periods 9 and 10, but in equilibrium, the investor may invest in both periods.

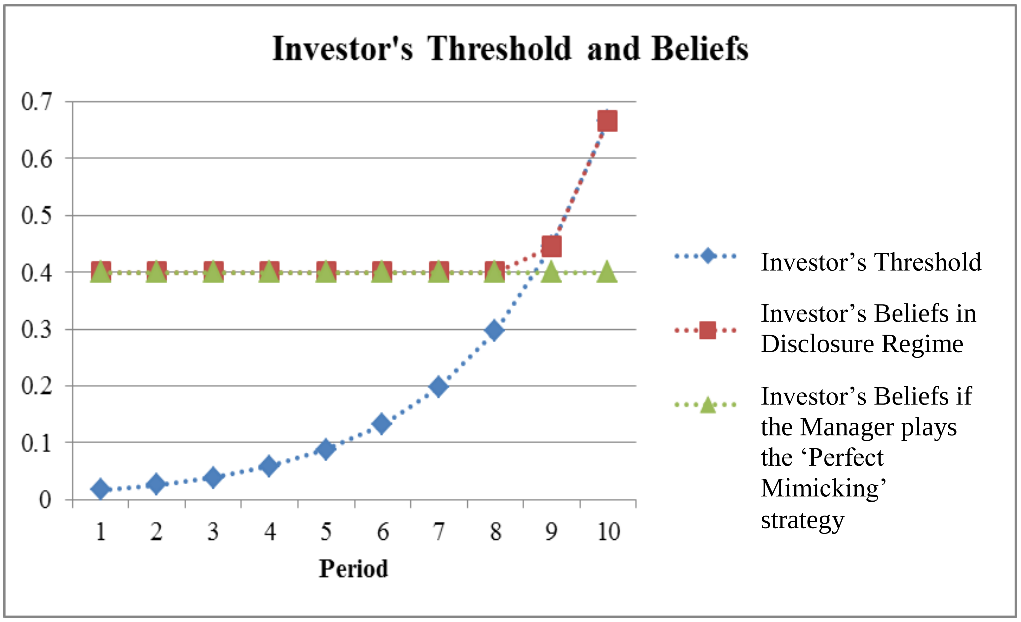

Figure 2B. *Investment threshold at which the investor invests with some probability and the investor's beliefs about the manager's type when n = 10, l = 1 , h = 5, and* P_0^D *= 0.4*

4.3. Properties of δ and *n*

From the investor's strategy outlined in the previous section, the investor chooses to invest in period 1 if $P_1^D = P_0^D = \delta > (4/(l+h))^n$. δ denotes the prior beliefs with which the game begins. It is the investor's ex ante beliefs about the manager's trustworthiness. $(4/(l+h))^n$ is the investment threshold for period 1. If δ is higher than this threshold, then the investor invests in period 1.

Note that as n increases, $(4/(l+h))^n$ decreases; that is, as the game is repeated over a larger number of periods, the investment threshold for period 1 becomes lower. With a large n, δ can be really small, and yet investment will occur in period 1.

4.4. Out-of-Equilibrium Beliefs

What happens if $P_t^D < (4/(l+h))^{n-t+1}$ but the investor chooses $m_t = e$, that is, the beliefs are below the investment threshold for period t but the investor chooses to invest? This is a violation of the investor's strategy (I(i)–I(iii)) defined earlier. The question is, what will his beliefs be in period $t + 1$ (denoted by P_{t+1}^D), given that he violates the equilibrium strategy in period t? To answer this, note first that $P_t^D < (4/(l+h))^{n-t+1} < (4/(l+h))^{n-t}$. Then, by M(iv), the rational manager chooses to mimic with probability S_t^D. Now, if $P_t^D > 0$, then by P(iii), $P_{t+1}^D = (4/(l+h))^{n-t}$, and if $P_t^D = 0$, then by P(v), $P_{t+1}^D = 0$.

Now, consider what happens if $P_t^D > (4/(l+h))^{n-t+1}$ but the investor chooses $m_t = 0$; that is, the beliefs are above the investment threshold for period t but the investor chooses not to invest. This is also a violation of the investor's strategy defined earlier (I(i)–I(iii)). Once again, the question arises, what will be the investor's beliefs in period $t + 1$ (denoted by P_{t+1}^D), given that he violates the equilibrium strategy in period t? By P(ii), $P_{t+1}^D = P_t^D$.

The rational manager can violate the equilibrium strategy by mimicking with less than certainty when he should be mimicking with certainty. This will lead to instances in which he either does not disclose or does not return (as dividend) to the investor half of what he receives as income of the firm. In either case, he will have revealed that he is untrustworthy/rational and, by P(iv), $P_{t+1}^D = 0$.

4.5. Comparison with Exogenously Imposed Information Sharing

Now consider the same game with the following modification: the investor learns λ_t; that is, the income of the firm is common knowledge between the investor and the manager. Consequently, the manager's disclosure decision is moot here, and a trustworthy manager is one who always chooses to return half of what he receives. The modified timeline is described in Figure 3.

Subject to the differences in how a trustworthy manager is defined, the equilibrium for this modified game with exogenously imposed information sharing is identical to that of the model described in Section 2, where information sharing was a choice of the manager. The investor continues to have a threshold strategy whereby he invests if his beliefs about the investor's trustworthiness are above the investment threshold and does not invest if his beliefs are below the threshold. On the threshold, the investor invests with a probability that makes the rational manager indifferent between mimicking and not mimicking to be a trustworthy type. The investment threshold for this modified game is the same as the investment threshold for the model from Section 2. (It is equal to $(4/(l+h))^{n-t+1}$.) Additionally, the probability with which the investor invests when his beliefs are exactly on the threshold is also the same.

The rational manager continues to mimic the behaviour of the trustworthy type to earn a reputation for being trustworthy. When the investor's beliefs about the manager's trustworthiness are above the threshold at which he will invest in the next period, then the rational manager mimics with certainty. When the said beliefs are below the said threshold, the rational manager switches to mimicking with a probability strictly less than 1. This probability is such that it makes the investor indifferent between investing and not investing and is equal to the corresponding probability from the model described in Section 2.

As with the model in Section 2, the game starts with pure strategy play and then switches to mixed strategy play. The point at which the switch occurs is identical, and as before, there is one exception—when the investor's ex ante beliefs about the manager's trustworthiness are above the investment threshold for the last period, then the game continues in pure strategies, and the switch does not occur.

The equilibrium in the case in which information sharing is exogenously imposed is identical to the case in which information sharing is voluntary because in the latter case, the rational manager (in equilibrium) always chooses to disclose up until such point as he does not pay back a dividend and thereby reveals that he is untrustworthy. Once he has revealed that he is untrustworthy/rational, it does not matter whether he chooses to voluntarily disclose his information. Note that when information sharing is exogenously imposed, dividend payment is the only tool available for reputation building, whereas when information is a choice of the manager, then dividend payment in conjunction with disclosure is used as a reputation-building tool. What is important is the role that information, whether exogenously imposed or voluntarily disclosed, plays in managerial efforts at building a reputation for being trustworthy. It reveals the state of nature to the investor and forces the rational manager to build a reputation along the lines outlined earlier. In the process, the investor receives a fair dividend from the manager.

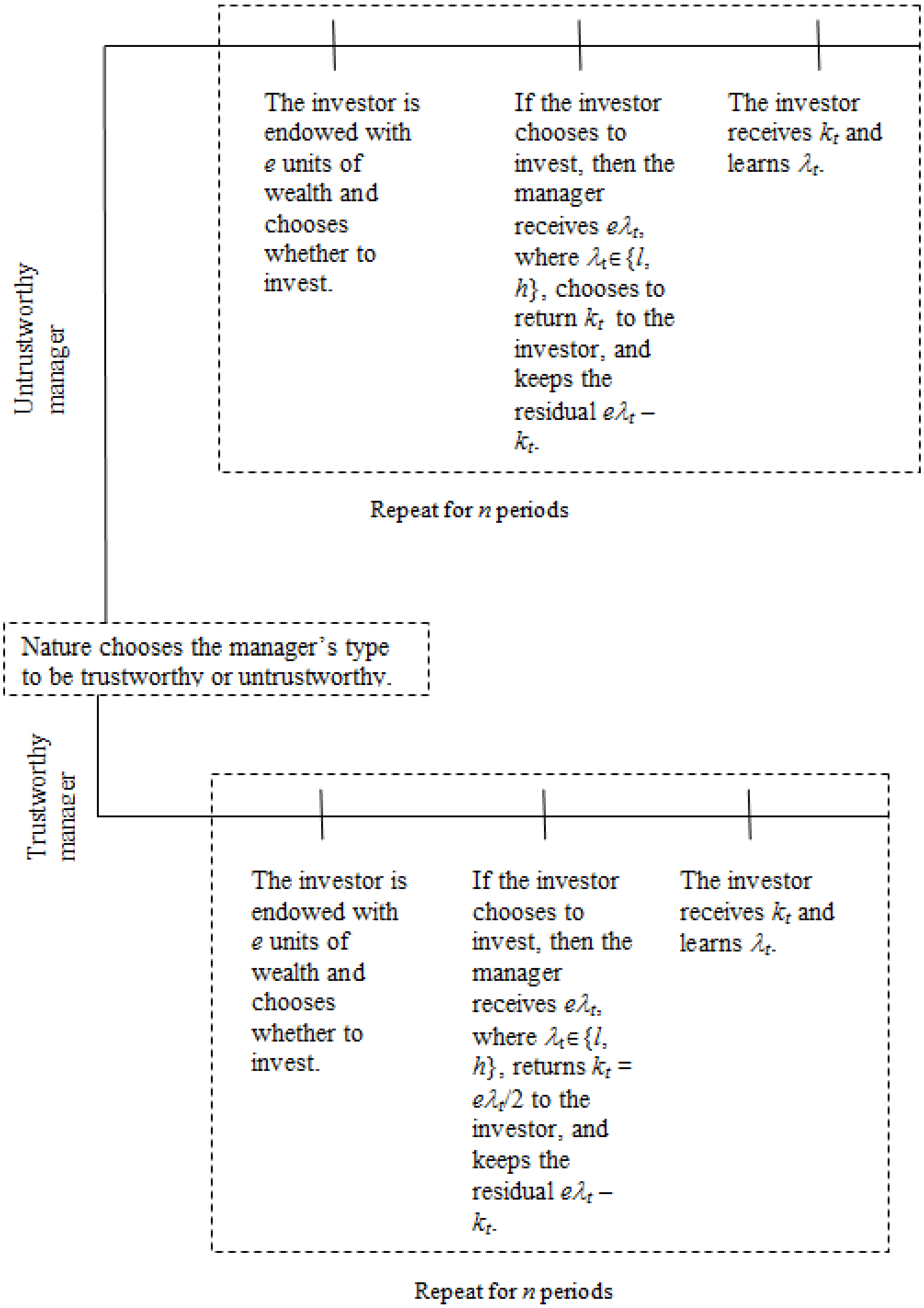

Figure 3. *Timeline of the modified game*

5. Conclusion

This article examines the role information plays in the building of trust and trustworthiness in complex economic settings in which there is separation of ownership and control of key economic resources. The investment game is a parsimonious way to capture the separation of ownership and management that has fuelled business and economic growth over the last several centuries. Modelling this gains further preeminence in today's economy, in which asset securitizations and derivative contracts epitomize such separation between risk bearers/owners and risk managers. Introducing disclosure in such a model provides a framework for analysis of the role of accounting information in a business superstructure built with the separation of risk bearing and management as one of its founding pillars. The agents in this game are strategic players, unlike the atomistic players of a rational expectations world. This mirrors more closely an economic world with a few big players, such as the one that faces us today, with few "too big to fail" economic agents.

In a setting with trustworthy and rational managers, choosing to voluntarily disclose private information is a natural act of the trustworthy manager, which the rational manager mimics to receive additional future investments. While anecdotally, it seems that information should provide opportunities for reputation building, there are subtleties in the process of reputation building that are not so obvious. For example, I show that mimicking with certainty is suboptimal and that a rational manager (except under certain specific circumstances) chooses mimicking with a certain probability strictly less than 1. The only exception is when the investor's beliefs about a manager's trustworthiness are exceptionally high; then, the manager chooses mimicking with certainty to maintain the beliefs at that high level.

When voluntary disclosure of private information is replaced by exogenously imposed information sharing, it results in a comparable equilibrium characterized by selective mimicking of the trustworthy type and thereby provides similar opportunities at facilitating the building of reputation for trust and trustworthiness. The setting with exogenously imposed information sharing may be thought of as a mandatory disclosure regime that obliterates information asymmetry by fiat. The results derived here then say that voluntary disclosure and mandatory disclosure are equivalent from a reputation-building perspective. This in turn implies that voluntary disclosure makes up for deficiencies in a reporting regime's mandatory requirements by providing as much opportunity for building trust and trustworthiness as exists in a regime where such disclosure is mandated. This explains a managerial incentive at providing information voluntarily, even when not required to.

Traditionally, the demand for accounting information is understood to arise from information needs of share markets or debt contracting (e.g., Ball, 2001; Ball *et al.*, 2008). Other literature (e.g., Grossman, 1981; Dye, 1985) bases the existence of voluntary disclosure of private information on investors' beliefs regarding managers' information about the value of the firm. Generally, managers with superior values voluntarily reveal their private information about value because otherwise, investors will assume that nondisclosed values are low. This article considers a model of reputation formation under different institutional structures of disclosure to show that voluntary disclosure of private information occurs because rational self-interested agents have an incentive to look like altruistic or trustworthy agents.

An interesting question is to look at what happens if the ability for voluntary disclosure of private information is removed from the disclosure regime. Defining such a regime and comparing it with the model here allows analysis of disclosure's reputation-building role from another perspective (Lunawat, 2012a). Disclosure is shown to occur in this article because of reputation building by rational agents. The effect of such reputation building on economic activity is an important question that calls for further study.

Given the complicated nature of the equilibrium described here, it is natural to ask if people actually behave as predicted, and this calls for an empirical examination (Lunawat, 2012b). While disclosure has been argued to be a managerial talent-signalling device, this article abstracts away from this question to focus on reputation building. An unanswered question, then, is the role reputation building may play in situations in which managers have different abilities in that a better manager has a higher probability of obtaining high firm productivity.

References

Ball, R. (2001) 'Infrastructure requirements for an economically efficient system of public financial reporting and disclosure', *Brookings–Wharton Papers on Financial Services.*

Ball, R., Robin, A. and Sadka, G. (2008) 'Is financial reporting shaped by equity markets or by debt markets? An international study of timeliness and conservatism', *Review of Accounting Studies*, 13, pp. 168–205.

Basu, S. and Waymire, G. (2006) 'Recordkeeping and human evolution', *Accounting Horizons*, 20(3), pp. 201–29.

Basu, S., Dickhaut, J., Hecht, G., Towry, K. and Waymire, G. (2009) 'Recordkeeping alters economic history by promoting reciprocity', *Proceedings of the National Academy of Sciences of the United States of America*, 106, pp. 1009–14.

Berg, J., Dickhaut, J. and McCabe, K. (1995) 'Trust, reciprocity and social history', *Games and Economic Behavior*, 10(1), pp. 122–142.

Beyer, A. and Dye, R. (forthcoming) ‘Reputation management and the disclosure of earnings forecasts’, *Review of Accounting Studies*.

Dickhaut, J., Lunawat, R., Waymire, G. and Xin, B. (2012) ‘Language, deception and social loss in an investment/trust game’, working paper, Emory University, University of Toronto. and University of St Thomas.

Dickhaut, J., Hubbard, J., Lunawat, R. and McCabe, K. (forthcoming) ‘Trust, reciprocity and interpersonal history: Fool me once, shame on you, fool me twice, shame on me’, *Journal of Economic Behavior and Organization*.

Dye, R. (1985) ‘Disclosure of nonproprietary information’, *Journal of Accounting Research,* 23, pp. 123–45.

Einhorn, E. and Ziv, A. (2008) ‘Intertemporal dynamics of corporate voluntary disclosures’, *Journal of Accounting Research,* 46, pp. 567–89.

Fudenberg, D. and Tirole, J. (1991) ‘Perfect Bayesian equilibrium and sequential equilibrium’, *Journal of Economic Theory,* 53(2), 236–60.

Grossman, S. (1981) ‘The informational role of warranties and private disclosure about product quality’, *Journal of Law and Economics*, 24, pp. 461–84.

King, R. (1996) ‘Reputation formation for reliable reporting: An experimental investigation’, *The Accounting Review,* 71(3), pp. 375–96.

Kreps, D. M. and Wilson, R. (1982) ‘Reputation and imperfect information’, *Journal of Economic Theory*, 27, pp. 253–79.

Kreps, D. M., Milgrom, P., Roberts, J. and Wilson, R. (1982) ‘Rational cooperation in the finitely repeated prisoners’ dilemma’, *Journal of Economic Theory,* 27, pp. 245–52.

Lunawat, R. (2012a) ‘Reputation effects of disclosure’, working paper, University of St Thomas.

Lunawat, R. (2012b) ‘An experimental investigation of reputation effects of disclosure in an investment/trust game’, working paper, University of St Thomas.

Appendix: Proof of Proposition

Proof. There are two things to verify. First, the investor’s beliefs must be consistent with the manager’s strategy in the sense that Bayes rule holds wherever applicable. Second, starting from any information set in the game, no player has a profitable deviation, that is, no player has an incentive to deviate.

I first verify the criterion of Bayesian consistency. If the manager chooses $d_1 = 0$, then it must be the case that the manager is untrustworthy and $P_{t+1}^D = 0$. If the investor does not invest in period t, then he does not learn anything about the manager’s type, and therefore $P_{t+1}^D = P_t^D$. If $P_t^D \geq (4/(l+h))^{n-t}$, the untrustworthy manager chooses $d_{t+1} = 1$ and $k_t = e\lambda_t / 2$. If $P_t^D = 0$, the untrustworthy manager chooses $k_t = 0$. In both cases, Bayes rule implies $P_{t+1}^D = P_t^D$. If $P_t^D \in (0, (4/(l+h))^{n-t})$, then with a probability S_t^D, the manager chooses $d_{t+1} = 1$ and $k_t = e\lambda_t / 2$, and with a probability $1 - S_t^D$, the manager chooses $k_t = 0$. In the instance the manager chooses $k_t = 0$, it must be the case that the manager is untrustworthy and $P_{t+1}^D = 0$. In the instance the manager returns half of what he receives in period t, Bayes rule requires the following:

$$P_{t+1}^D = \text{Prob (the manager is trustworthy | the manager returns half of what he receives)}$$

$$= \frac{\text{Prob (the manager is trustworthy and returns half of what he receives)}}{\text{Prob (the manager returns half of what he receives)}}$$

$$= \frac{\text{Prob (the manager returns half |the manager is trustworthy) Prob (the manager is trustworthy)}}{\left(\begin{array}{l}\text{Prob (the manager returns half | the manager is trustworthy) Prob (the manager is trustworthy)} \\ \text{+Prob (the manager returns half | the manager is untrustworthy) Prob (the manager is untrustworthy)}\end{array}\right)}$$

$$= \frac{1.P_t^D}{1.P_t^D + S_t^D(1-P_t^D)}$$

$$= \left(\frac{4}{l+h}\right)^{n-t}$$

This satisfies the criterion of Bayesian consistency.

I next verify that the investor's strategy is optimal. The investor's payoff from not investing (choosing $m_t = 0$) = e. If $P_t^D \geq (4/(l+h))^{n-t}$, the untrustworthy manager chooses $k_t = e\lambda_t / 2$, and then the investor's expected payoff from investing (choosing $m_t = e$) $= (e/2)((l+h)/2)$. Because $(l + h) > 4$, the investor's expected payoff from choosing $m_t = e$ is greater than the investor's payoff from choosing $m_t = 0$, making it optimal for the investor to choose $m_t = e$.

If $P_t^D < (4/(l+h))^{n-t}$, then with probability S_t^D, the untrustworthy manager chooses $k_t = e\lambda_t / 2$, and with probability $1 - S_t^D$, the untrustworthy manager chooses $k_t = 0$. Therefore the investor's expected payoff from investing (choosing $m_t = e$) $= e/2((l+h)/2)(P_t^D + (1-P_t^D)S_t^D)$.

Inserting $S_t^D = P_t^D(1-(4/(l+h))^{n-t})/(4/(l+h))^{n-t}(1-P_t^D)$ in the expression for the investor's expected payoff and simplifying the expression yields the investor's expected payoff $= eP_t^D((l+h)/4)^{n-t+1}$. This implies that the investor's expected payoff from choosing $m_t = e$ is greater when $P_t^D > (4/(l+h))^{n-t+1}$ and the investor's expected payoff from choosing $m_t = 0$ is greater when $P_t^D < (4/(l+h))^{n-t+1}$. (Note that $P_t^D < (4/(l+h))^{n-t+1} < (4/(l+h))^{n-t}$.) If $P_t^D = (4/(l+h))^{n-t+1}$, then the investor's expected payoff from choosing $m_t = e$ is equal to the investor's payoff from choosing $m_t = 0$. At this point, the investor is indifferent between choosing $m_t = e$ and choosing $m_t = 0$. Therefore, with probability V_t^D, the investor chooses $m_t = e$, and with probability $1-V_t^D$, the investor chooses $m_t = 0$. The choice of the probability V_t^D is such that it makes the rational manager indifferent between choosing $k_{t-1} = e\lambda_{t-1}/2$ and choosing $k_{t-1} = 0$. To verify the manager's indifference at this point, note that the manager's payoff from choosing $k_{t-1} = 0$ is $e\lambda_{t-1}$ and the manager's expected payoff from choosing $k_{t-1} = e\lambda_{t-1}/2$ is $e\lambda_{t-1}/2 + eV_t^D(l+h)/2$. Because $V_t^D = \lambda_{t-1}/(l+h)$, the manager's expected payoff from choosing $k_{t-1} = 0$ equals the manager's payoff from choosing $k_{t-1} = e\lambda_{t-1}/2$.

Finally, I verify that the untrustworthy/rational manager's strategy is optimal. If $d_1 = 0$, it implies that the manager must be untrustworthy ($P_1^D = 0$), and therefore the investor chooses $m_1 = 0$. If $d_1 = 1$, it implies that the manager may be trustworthy ($P_1^D = P_0^D = \delta$), and then if $\delta > (4/(l+h))^n$, the investor will invest[2]. Consequently, it is optimal for the manager to choose $d_1 = 1$. If $t = n$, the manager's payoff from choosing $k_n = 0$ is $e\lambda_n$, while the manager's payoff from choosing $k_n = e\lambda_n/2$ is $e\lambda_n/2$, making it optimal for the manager to choose $k_n = 0$.

If $t < n$ and $P_t^D > (4/(l+h))^{n-t}$, then the investor chooses $m_{t+1} = e$. Thus the manager's payoff from choosing $k_t = 0$ is $e\lambda_t$, while the manager's expected payoff from choosing $k_t = e\lambda_t/2$ is $e\lambda_t/2 + eE(\lambda_{t+1}) = e\lambda_t/2 + e(l+h)/2$, making it optimal for the manager to

[2] If $\delta \le (4/(l+h))^n$, the investor will not invest, making the rational manager indifferent between $d_1 = 0$ and $d_1 = 1$.

choose $k_t = e\lambda_t / 2$. If $P_t^D = (4/(l+h))^{n-t}$, then the investor chooses m_{t+1} = e with probability V_{t+1}^D, and the manager's expected payoff from choosing $k_t = e\lambda_t / 2$ is $e\lambda_t / 2 + eE(\lambda_{t+1})V_{t+1}^D = e\lambda_t / 2 + e((l+h)/2)(\lambda_t/(l+h)) = e\lambda_t$. At this point, the manager is indifferent between choosing $k_t = 0$ and choosing $k_t = e\lambda_t / 2$. Therefore the manager chooses $k_t = 0$ with such a probability (and $k_t = e\lambda_t / 2$ with complementary probability) that makes the investor indifferent between choosing m_{t+1} = e and choosing m_{t+1} = 0. The investor is indifferent between choosing m_{t+1} = e and choosing m_{t+1} = 0 when $P_{t+1}^D = (4/(l+h))^{n-t}$. Now, $P_{t+1}^D = P_t^D = (4/(l+h))^{n-t}$ requires the manager choose $k_t = e\lambda_t / 2$ with probability 1.

Now I am required to verify that if t < n and $P_t^D < (4/(l+h))^{n-t}$, then with a probability $S_t^D = P_t^D(1-(4/(l+h))^{n-t})/(4/(l+h))^{n-t}(1-P_t^D)$, the manager chooses to return half of what he receives in period t. Suppose instead that the manager chooses $k_t = e\lambda_t / 2$ with a probability $S_t^D + \varepsilon$ for some $\varepsilon > 0$ such that $S_t^D + \varepsilon \le 1$. Then, the posterior probability P'_{t+1} is given by $P_t^D / \left(P_t^D + (S_t^D + \varepsilon)(1-P_t^D)\right)$. Now, $P'_{t+1} < P_{t+1}^D = (4/(l+h))^{n-t}$. Therefore, the investor chooses m_{t+1} = 0. This implies that the sum of the manager's expected payoff in periods t and (t + 1) = $e\lambda_t(1-S_t^D-\varepsilon) + e(\lambda_t/2)(S_t^D+\varepsilon)$. In contrast, if the manager chooses $k_t = \lambda_t / 2$ with a probability S_t^D, then his expected payoff = $E_t\left[e\lambda_t(1-S_t^D) + S_t^D(e\lambda/2_t + eV_{t+1}^D\lambda_{t+1})\right] = e\lambda_t$. Because the manager's expected payoff from choosing $k_t = e\lambda_t / 2$ with a probability S_t^D is higher than his expected payoff from choosing $k_t = e\lambda_t / 2$ with a probability $S_t^D + \varepsilon$, he does not choose $k_t = e\lambda_t / 2$ with a probability $S_t^D + \varepsilon$. Similarly, it can be shown that the manager does not choose $k_t = e\lambda_t / 2$ with a probability $S_t^D - \varepsilon$ for some $\varepsilon > 0$ such that $S_t^D - \varepsilon > 0$. The only case remaining to be ruled out is one in which the manager chooses $k_t = e\lambda_t / 2$ with a probability $S_t^D - \varepsilon$ for some $\varepsilon > 0$ such that $S_t^D - \varepsilon = 0$, that is, the case in which the manager never chooses $k_t = e\lambda_t / 2$. In this case, the posterior probability $P''_{t+1} = 1$; that is, given a return of $k_t = e\lambda_t / 2$, the investor believes with probability 1 that the manager is trustworthy. This strategy can never constitute an equilibrium because a profitable deviation for the untrustworthy manager is to mimic to be the trustworthy type and choose $k_t = e\lambda_t / 2$, and then because $P''_{t+1} = 1$, the investor will chose $m_{t+1} = e$.

This completes the proof that the set of beliefs and strategies described earlier constitutes a sequential equilibrium. Subject to the arbitrariness involved in specifying the out-of-equilibrium beliefs in a sequential equilibrium, the equilibrium described here is unique, and the uniqueness follows from the proof outlined in this appendix.